DOCUMENT RESUME

ED 441 486 IR 057 757

AUTHOR	Place, Emma
TITLE	International Collaboration on Internet Subject Gateways.
PUB DATE	1999-08-00
NOTE	9p.; In: IFLA Council and General Conference. Conference Programme and Proceedings (65th, Bangkok, Thailand, August 20-28, 1999); see IR 057 674.
AVAILABLE FROM	For full text: http://www.ifla.org/IV/ifla65/papers/009-143e.htm.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Directories; Foreign Countries; *Gateway Systems; Guidelines; *Information Services; *Internet; Library Role; Online Searching; Standards
IDENTIFIERS	Europe; International Standards; Search Engines; Web Sites

ABSTRACT

A number of libraries in Europe are involved in the development of Internet subject gateways--services that aim to help users find high quality resources on the Internet. Subject gateways such as SOSIG (Social Science Information Gateway) have been available on the Internet for some years now, and they offer an alternative to Internet search engines such as AltaVista and to directories such as Yahoo. Distinctively, subject gateways draw upon the skills, practices, and standards of the international library community and apply these to Internet-based information. This paper suggests that librarians are ideally placed to play a major role in building Internet resource discovery services and that subject gateways offer a means to do this. It outlines some of the subject gateway initiatives in Europe and describes the tools and technologies developed by the DESIRE project to support the development of new gateways in other countries. It also discusses how IMesh, a group for gateways from around the world, aims to work on an international strategy for subject gateways and on developing standards to support this. Contains 22 Web resources. (Author/MES)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 486

65th IFLA Council and General Conference

Bangkok, Thailand, August 20 - August 28, 1999

Conference Proceedings

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
A.L. Van Wesermael
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

Code Number: 009-143-E
Division Number: VI
Professional Group: Information Technology
Joint Meeting with: -
Meeting Number: 143
Simultaneous Interpretation: *Yes*

International Collaboration on Internet Subject Gateways

Emma Place (nee Worsfold)
Institute for Learning and Research Technology
University of Bristol, UK

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Abstract

A number of libraries in Europe are involved in the development of Internet subject gateways - services that aim to help users find high quality resources on Internet. Subject gateways such as SOSIG (The Social Science Information Gateway) have been available on the Internet for some years now, and they offer an alternative to Internet search engines such as AltaVista and to directories such as Yahoo . Distinctively, subject gateways draw upon the skills, practices and standards of the international library community and apply these to Internet-based information. This paper will suggest that librarians are ideally placed to play a major role in building Internet resource discovery services and that subject gateways offer a means to do this. It will outline some of the subject gateway initiatives in Europe and will describe the tools and technologies developed by the DESIRE project to support the development of new gateways in other countries. It will also discuss how IMesh , a group for gateways from around the World, aims to work on an international strategy for subject gateways and on developing standards to support this.

Paper

Background

> ***"The Web is quickly becoming the world's fastest growing repository of data." "***
> (Tim Berners-Lee, W3C director and creator of the World Wide Web (WWW)

This is a time of upheaval for the library profession, as the Internet becomes a major medium

IR057757
BEST COPY AVAILABLE

in the information world. The Internet offers access to myriad information resources but the fact remains that it is still very hard for people to locate high quality information amid the general chaos. In the past few years the issue of resource discovery on the Internet has been the focus of much work by many different communities.

Search Engines

The Internet Search engines, such as AltaVista, and Excite, rely on automated solutions to resource discovery. They send out robots or Web crawlers to trawl the Internet and automatically index the files that they find. These indexes can then be searched by keyword and return records which contain automatically generated descriptions of the resources, usually the first few paragraphs of the resource itself. Search engines are good for finding lots of information - a search often yields thousands of resources. However, the results can be overwhelming, unmanageable, full of irrelevant references and are often too prolific to meet user needs.

Web Directories

Web directories such as Yahoo and The Open Directory are, in a sense the Internet equivalent of a public library that is not staffed by librarians! They rely on human input to create directories on the WWW that list Internet resources, with each one described briefly and classified under a subject heading. These directories aim to describe large numbers of Internet resources and include both serious and recreational sites.

The Open Directory is a remarkable project, since, in a sense, the general public are invited to build their own library on the Internet - selecting, classifying and "cataloguing" resources. The Open Directory has a volunteer work force of Editors (currently over 6,000 of them), who spend time adding resources and resource-descriptions to the directory (currently over 100,000!). Both Yahoo and The Open Directory aim to be the biggest Internet directories, with a high level of coverage and popular appeal as high priorities.

Internet Subject Gateways

Subject gateways offer an alternative to the Internet search engines and Web directories. What is the definition of a subject gateway? In some sense they are the Internet equivalent of an academic or special library. Subject gateways are Internet-based services designed to help users locate high quality information that is available on the Internet. They are typically, a database of detailed metadata (or catalogue) records which describe Internet resources and offer a hyperlink to the resources. Users can chose to either search the database by keyword, or to browse the resources under subject headings. Subject gateways are characterised by two key factors:

- They are quality controlled
- They are built by subject and information specialists - often librarians

Quality Selection Procedures

Formal quality selection criteria are used to guide collection development within the gateways. Examples of the selection policies of gateways have been collected by the DESIRE project . DESIRE has also produced an online tutorial called "Internet Detective" that aims to teach the skills required to evaluate the quality of resources on the Internet and this gives some insight into the sort of work that gateway staff do in evaluating and selecting Internet resources.

Classification of Internet Resources

Classification schemes are used by gateways to set up the browsing option for users. Many

gateways use traditional library classification schemes such as Dewey Decimal classification or Universal Decimal classification. A report on the use of classification schemes in Internet services has been produced by the DESIRE project, which describes this usage in more detail .

Standard Metadata Formats

Standard metadata formats are used when describing an Internet resource in a database record. These formats support effective information retrieval from the databases, but also ensure that gateways can inter-operate with eachother and, potentially, with other databases such as library OPACS. These standards also give the option of converting and mapping one format to another, which could be important as Web metadata standards develop and change. In 1997 DESIRE produced a comprehensive review of metadata formats . In the UK, UKOLN (The UK Office of Library Networking) has a Metadata Group that conducts ongoing research into metadata formats, especially in relation to library cataloguing formats such as MARC. Their Web site offers software tools for handling metadata and information on mapping between metadata formats .

European Gateway Initiatives

A number of Internet subject gateways have been developed in Europe and a significant community of libraries involved in gateways is developing.

- United Kingdom - The Resource Discovery Network

  In the UK a number of subject gateways are being funded by the UK government's Higher Education Funding Council and are organised under the Resource Discovery Network (RDN). All the UK gateways are based in universities and involve input from librarians and information professionals:

  - SOSIG (Social Science)
  - EEVL (Engineering)
  - OMNI (Biomedical information)
  - Biz/ed (Business and Economics Education)
  - History
  - ADAM (Art, Design, Architecture & Media)

- · The Netherlands - DutchESS

  The National Library of the Netherlands (Koninklijke Bibliotheek) has built a subject gateway in co-operation with seven university libraries called DutchESS (Dutch Electronic Subject Service) - a national gateway, covering all subjects.

- Finland - The Finnish Virtual Library Project

  In Finland the government's Ministry of Education has funded the large-scale development of national subject gateways. The Finnish Virtual Library project was launched in 1995 and involves collaborative work between eight university libraries.

- Sweden - EELS

  EELS covers the broad subject area of Electronic Engineering. It is a cooperative project of the six Swedish University of Technology Libraries.

- Denmark (and other Nordic Countries) - NOVAGate

  NOVAGate covers forestry, veterinary, agricultural, food and environmental sciences

and is produced by the libraries of the NOVA University in Denmark, Finland, Iceland, Norway and Sweden.

The DESIRE Project

DESIRE is an international project funded by the European Union. The project aims to facilitate use of the World Wide Web among Europe's research community and one of the ways it is doing this is by developing and promoting the Internet subject gateways model. SOSIG, DutchESS and EELS are all partners in the DESIRE project and have been working with other gateways (including the Finnish Virtual Libraries Project and NOVAGate).

DESIRE Workshop for Europe's National Libraries

There is considerable scope for the library community to be involved in Internet subject gateways. As illustrated in the gateways described above, many libraries in many countries are already seeing work on gateways as an important part of their remit. Once a country has a gateway structure in place, librarians from across that country can work collaboratively to build the collection. The subject gateways model offers strategic and standardardised methods for doing this. DESIRE aims to support the development of new gateways in Europe, especially large-scale national gateways. In September 1999 there will be a DESIRE workshop for the National libraries of Europe : ***"Building national and large-scale Internet Information Gateways: a workshop for the National Libraries of Europe"*** At the time of writing, seventeen European National libraries have signed up for the DESIRE workshop, and together we hope to make some important steps towards building a European network of gateways.

As the Internet continues to expand so quickly it is clear that no single gateway or country can hope to catalogue all the Internet resources available. A distributed model is required, where each country takes responsibility for describing the high quality resources available on its national network. Imagine the scenario where librarians from every country work at building a gateway to the best of their national Internet resources. Imagine then, that it is possible to cross-search any combination of these gateways - to find high quality Internet resources from around the World. In fact, the technologies and standards already exist to make this vision a reality. What still requires a lot of work is the development of the human networks that can maximise the potential of these standards and technologies - and the library community is perfectly placed to take up this challenge! Building an international network of gateways takes time, but the library community has both the expertise and the commitment to develop these valuable Internet search tools.

Distributed Teams of Librarians

Subject gateways provide a successful model for involving the library community in Internet resource discovery. Existing gateways have invested effort in developing systems that support the work of distributed teams, so that a librarian can work on a gateway from anywhere in the World as long as they have access to a networked PC and a Web browser. Distributed Internet cataloguing means that libraries can contribute to a shared service, rather than having to each build a local service. This is an efficient way of working - it avoids duplicated effort and collaboration means large-scale gateways with much better coverage can be developed. Many of the gateways described above benefit from the input of a distributed team of librarians. A DESIRE report "Distributed and Part-Automated Cataloguing" describes the different models being used by existing gateways. The ROADS software supports distributed cataloguing by providing a Web interface to the database. Records can be added, deleted or edited remotely. All this work can be done via the Web - the teams can work from their own offices using their own workstations and fit this "Internet librarianship" in alongside their usual work in the library.

Distributed Databases

The technologies also exist to support cross-searching of distributed databases. Interoperability has been the focus of much research by DESIRE and ROADS and other communities. If different databases of metadata records can be cross-searched this offers the potential for different communities to work at describing different sections of the Internet and for end-users to cross-search all these collections simultaneously. On a national level both the UK gateways and the Finnish Virtual library project are working on cross-searching distributed gateway databases. The end-user remains blissfully unaware of the complex organisation behind their search - from their point of view they are making a single search from a single Web page and get a single page of results.

SOSIG and Biz/ed have already implemented cross-searching into their working services. When users search SOSIG they are in fact, also cross-searching the Biz/ed database - results from the two databases are returned on the same page. The technologies used to achieve this are described in a paper published in Dlib magazine . Databases located in different countries can also be cross-searched simultaneously - DutchESS (in the Netherlands) has been working closely with SOSIG (in the UK) to set up a cross-search mechanism, so that both the collections can be accessed simultaneously by users from both countries (and indeed elsewhere!). This is pioneering work and when it is in place, it is hoped the same mechanism will be used by other gateways to set up similar systems. Demonstrations of the cross-searching work being done by DESIRE and ROADS is available on the DESIRE Web site

Tools for Building Large-Scale Internet Subject Gateways

DESIRE is developing tools and methods for the development of large-scale Internet subject gateways. It is also working with both library and Internet standards organisations to develop standard practices for developing gateways, to ensure that they are interoperable and can work together to form large-scale, collaborative services.

The DESIRE Gateways Handbook

Later this year DESIRE will publish the "Information Gateways Handbook" - a guide for libraries interested in setting up large-scale subject gateways of their own. The Handbook will be made freely available on the WWW and will describe all the methods and tools required to set up a large-scale Internet subject gateway. It draws upon over three years of DESIRE research into subject gateways and will include case studies and examples from many of the gateways described earlier in this paper. It is hoped that the Handbook will assist other countries to set up their own national gateway initiatives so that more libraries and more librarians can begin to play a role in Internet resource discovery.

ROADS

ROADS is a set of software tools to enable the set up and maintenance of Web based subject gateways. It was developed as part of the UK's Electronic Libraries Programme but is freely available for anyone to use. The software includes the database technology required to set up a gateway, the administration centre required to facilitate remote cataloguing via the WWW and everything else needed to run a gateway. Many of the gateways described above use ROADS, notably SOSIG and The Finnish Virtual Library Project. The ROADS open-source software toolkit is being produced by a consortium of developers with expertise in network-based resource identification, indexing and cataloguing. This has resulted in a standards-based approach to software development, making it compatible with current, and developing indexing and cataloguing requirements. In addition, there is ample documentation and online support for people interested in using the software for either experimental purposes or service provision.

IMesh : the International Gateway Community

IMesh is a collaborative network, involving key players in the World's subject gateway community (not only those in Europe). It is likely that IMesh will be the key player in future gateway developments internationally.

IMesh was formed as a result of a meeting at The Second European Conference on Research and Advanced Technology for Digital Libraries, held in Crete in September 1998, attended by 25 delegates from 15 countries. One of the main aims of IMesh is to explore the potential for collaborative development of gateways internationally. It would require significant investment of effort and resources for a single country to attempt to create a gateway that pointed to the best of the Internet from all countries, in all languages in all subject areas. The IMesh group is looking at ways in which the effort can be shared through international collaborative agreements. Many of the technologies required for cross-searching different gateways and for remote cataloguing into gateways already exist. What is lacking is the strategic organisation between gateways and IMesh aims to address this. The report from this meeting outlines the main issues that the group aims to address. In June 1999 the first IMesh workshop will be held in Warwick in the UK and will be attended by gateway providers from around the World. I hope to report on the outcomes of this meeting when I present this paper at the IFLA conference. An IMesh discussion list exists and those interested in international collaboration amongst subject gateways are invited to join. The list provides an open forum for exchanging ideas and technology for promoting the subject gateway movement. Details are on the IMesh Web site .

Future and Conclusions

In many ways the Internet is still a bit of a building site! Many things are still under-construction, including the basic architecture of the Web. The World Wide Web Consortium is still working on building a structure that can support resource discovery on the Internet. They have recently released the Resource Description Framework (RDF) model and syntax specification, which aims to provide a basic infrastructure on the Web to support the transfer and processing of metadata. This marks a new age on the Web as in effect, it allows anyone to "catalogue" a Web resource in a machine understandable way. Different people will want to use RDF in different ways - it is simply the structure within which different people can work. Gateways are working with the W3C to see how RDF can support these high quality metadata collections. Potentially, librarians could forge the same role for themselves on the Internet that they have had traditionally - as third party information providers that end-users can learn to trust and rely on when searching for information.

Although the structure for the Internet library is not yet complete, it does not mean that librarians have to wait to start building their Internet collections. The human networks required to effectively catalogue the Internet will take many years to build. Libraries can work on creating metadata records for Internet resources and on finding their place in the metadata community. They can also start becoming familiar with the new metadata and Internet cataloguing standards so all the records are compatible. Although this paper is being presented in the "Information Technology Section" of IFLA, in many ways the technologies are the least of our worries - it is the human factor that now requires significant development.

DESIRE and IMesh hope to facilitate the involvement of the wider library community in Internet resource discovery. In this paper I have described services that already involve input from large numbers of libraries and librarians. Perhaps the IFLA community can help us to take this work forward and to promote Internet Librarianship as an important new role for the profession.

The Open Directory Project

http://dmoz.org/Computers/Internet/WWW/Directories/Open_Directory_Project/

Selection Criteria: Examples
http://www.desire.org/results/discovery/cat/selectex_des.htm

Internet Detective
http://www.sosig.ac.uk/desire/internet-detective.html

The Role of Classification Schemes in Internet Resource Description and Discovery
http://www.desire.org/results/discovery/cat/class_des.htm

A Review of Metadata: a Survey of Current Resource Description Formats
http://www.desire.org/results/discovery/cat/meta_des.htm

UKOLN Metadata Web Page
http://www.ukoln.ac.uk/metadata/

EEVL (Edinburgh Engineering Virtual Library)
http://www.eevl.ac.uk/

OMNI
http://omni.ac.uk/

Biz/ed
http://www.bized.ac.uk/

HISTORY
http://ihr.sas.ac.uk/

ADAM
http://www.adam.ac.uk/

DutchESS (Dutch Electronic Subject Service)
http://www.konbib.nl/dutchess/

The Finnish Virtual Library Project
http://www.uku.fi/kirjasto/virtuaalikirjasto/

EELS
http://www.ub2.lu.se/eel/eelhome.html

Building National and Large-scale Internet Information Gateways: a Workshop for the National Libraries of Europe (DESIRE)
http://www.desire.org/html/subjectgateways/workshops/workshop1.html

Distributed and Part-Automated Cataloguing (DESIRE)
http://www.sosig.ac.uk/desire/cat/cataloguing.html

Cross-Searching Subject Gateways: The Query Routing and Forward Knowledge Approach (D-Lib Magazine, January 1998)
http://www.dlib.org/dlib/january98/01kirriemuir.html

DESIRE Cross-searching Demonstrator
http://www.desire.org/ROADS/cgi-bin/search.pl

ROADS
http://www.ilrt.bris.ac.uk/roads/

D-Lib report from the Crete ECDL'98 Subject Gateways meeting
http://www.dlib.org/dlib/october98/10clips.html#GATEWAYS

IMesh: International Collaboration on Internet Subject Gateways
http://www.desire.org/html/subjectgateways/community/imesh/

W3C (World Wide Web Consortium)
http://www.w3.org/

Latest Revision: *May 27, 1999*

Copyright © 1995-1999
International Federation of Library Associations and Institutions
www.ifla.org